SPECIFICATION

TO ALL WHOM IT MAY CONCERN:

BE IT KNOWN THAT WE, Hitoshi Ozawa, a citizen of
Japan residing at Ota, Japan and Kyoko Kawazu, a citizen of
Japan residing at Ota, Japan have invented certain new and
useful improvements in

PROCESS OF AUTOMATICALLY GENERATING
TRANSLATION-EXAMPLE DICTIONARY, PROGRAM PRODUCT,
COMPUTER-READABLE RECORDING MEDIUM AND
APPARATUS FOR PERFORMING THEREOF

of which the following is a specification : -

TITLE OF THE INVENTION

PROCESS OF AUTOMATICALLY GENERATING TRANSLATION-EXAMPLE DICTIONARY, PROGRAM PRODUCT, COMPUTER-READABLE RECORDING MEDIUM AND APPARATUS FOR PERFORMING THEREOF

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to a machine-translating process and particularly relates to a process of automatically generating a translation-example dictionary in which some portions of translation-example information are expressed by variables.

2. Description of the Related Art

When translating an original text written in a source language into another language using a machine translation system, it is often insufficient to use a basic dictionary and/or a technical term dictionary provided in a computer for providing a translation which corresponds well to the meaning of the original text. Accordingly, an Example-based Machine Translation process has been proposed in which frequently-used translation examples each including an original text and a translated text are pre-registered by a person in charge of translation and then a translation process of the relevant original text is performed.

A translation-example dictionary used in the Example-based Machine Translation may be a type of dictionary in which some portions of the translation example information, for example nouns, are expressed as variables. Such a dictionary is used for a machine translation process in which, when there is a correspondence between the original text to be translated and the original text data in the translation-example dictionary where from the

variable have been excluded, the variables of the original text data and the translated text data are substituted by data included in the basic dictionary and/or technical term dictionary.

In order to create a translation-example dictionary having some portions of the translation-example information expressed as variables, the person in charge of translation must determine which portion of the original text and the translated text are to be expressed as variables. Also, since the variables must be manually registered into the dictionary by the person in charge of translation, many steps are required for creating the translation-example dictionary, and also, there is a problem that the judging criteria for portions to be expressed as variables may differ between different persons in charge of translation.

Also, it is understood that the reliability of a translated text obtained by translation using a translation-example dictionary having many variables is lower than the reliability of a translated text obtained by translation using a translation-example dictionary having a few variables. However, there is no material provided for determining the reliability of the translated text.

SUMMARY OF THE INVENTION

Accordingly, it is a general object of the present invention to provide a process for automatically generating a translation-example dictionary which can solve the problems described above.

It is another and more specific object of the present invention to provide a process for automatically generating a translation-example dictionary in which some portions of the

translation-example information are expressed as variables, and to record a number of portions expressed as variables in the translation-example information.

5   In order to achieve the above object, a process of a process of creating a translation-example dictionary for an Example-based Machine Translation is provided, the process including the steps of:

10   a) comparing first translation-example information and another first translation-example information to detect if there is any different portion;

      b) specifying a word class of each of
15 different portions, if any, detected in the step a);

      c) generating variables by linking the at least one different portion detected in the step a) and the word class specified in the step b) so as to create second translation-example information; and

20   d) registering the second translation-example information into the translation-example dictionary.

      The process of the present invention is particularly useful for automatically generating a
25 translation-example dictionary in which some portions of the translation-example information are expressed as variables. For example, the translation-example dictionary in which a part of the translation-example information are expressed as
30 variables can be automatically generated from an existing translation-example dictionary without needing many steps and in such a manner that judgment criteria for portions to be expressed as variables do not differ between different persons in
35 charge of translation.

      Also, as in a case where it is required to distinguish between a translated text with a

comparatively high reliability which has been translated using a translation-example information with less portions expressed as variables and a translated text with a comparatively low reliability
5   which has been translated using a translation-example information with more portions expressed as variables, and display them separately, a number of portions expressed as variables in the translation-example information may also be recorded.
10

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a schematic diagram showing an information processing apparatus which may implement a process of the present invention.
15   Fig. 2 is a system block diagram showing the construction of an important part within the main body part of the computer system of Fig. 1.

Fig. 3 is a schematic diagram showing a principle structure of the present invention.
20   Fig. 4 is a diagram showing a first translation-example dictionary.

Fig. 5 is a diagram showing a second translation-example dictionary.

Fig. 6 is a flow chart showing various
25   steps for reading out first translation-examples and a step of morphological analysis of the original text.

Fig. 7 is a flow chart showing various steps for determining whether the read out original
30   texts are similar.

Fig. 8 is a flow chart showing various steps for expressing the differing parts in the original text as variables and a step of morphological analysis of the translated text.
35   Fig. 9 is a flow chart showing various steps for expressing the differing parts in the translated text as variables.

Fig. 10 is a flow chart showing various steps for registering second translation-example information into the second translation-example dictionary.

Fig. 11 is a diagram showing a manner in which variables are generated.

Fig. 12 is a schematic diagram showing a process of creating the second translation-example information.

Fig. 13 is a schematic diagram showing a manner in which the translation results are color-coded in accordance with the number of variables.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

In the following, principles and embodiments of the present invention will be described with reference to the accompanying drawings.

Fig. 1 is a schematic diagram showing an information processing apparatus which may implement a process of the present invention. A computer system shown in Fig. 1 is formed by a general computer system such as a personal computer (PC). The computer system 10 generally includes a main body part 11 which includes a CPU, a disk drive and the like, a display unit 12 which displays an image on a display screen 12a in response to an instruction to the computer system 10, a keyboard 13 which is used to input various kinds of information to the computer system 10, a mouse 14 which is used to make access to an external data base or the like and to download a program or the like stored in another computer system. A program which is stored in a portable recording medium such as a disk 17 or is downloaded from a recording medium 16 of another computer system by use of a communication unit such as the modem 15, is input to and is compiled in the

computer system 10. This program includes a program for causing the CPU of the computer system 10 to carry out a process of creating a translation-example dictionary for using in an Example-based

5   Machine Translation.

Fig. 2 is a system block diagram showing the construction of an important part within the main body part 11 of the computer system 10 of Fig. 1. In Fig. 2, the main body part 11 generally

10   includes a CPU 21, a memory part 22 including a RAM, ROM or the like, a disk drive 23 for the disk 17, and a hard disk drive 24 which are coupled via a bus 25. The display unit 12 and the like are coupled to the bus 25.

15   The construction of the computer system 10 is not limited to that shown in Figs. 1 and 2 and various known constructions may be used in place thereof.

Referring now to Figs. 3 to 5, an

20   embodiment of the present invention will be described.

Fig. 3 is a schematic diagram showing a principle structure of the present invention. An automatic translation-example dictionary generating

25   program 100 includes a translation-example input step 101, a translation-example comparing step 102, a word class specifying step 103, a variable generating step 104 and a translation-example dictionary registering step 105.

30   An automatic translation-example dictionary generating apparatus 150 includes a first translation-example dictionary 110, the automatic translation-example dictionary generating program 100 which may operate in the apparatus 150, and a

35   second translation-example dictionary 120.

As shown in Fig. 4, the first translation-example dictionary 110 includes first translation-

example information 200 each having an original text data 201 and a translated text data 202 and not having any variable portions.

As shown in Fig. 5, the second translation-example dictionary 120 includes second translation-example information 300 each having an original text data 301 and a translated text data 302 and having at least one variable portion. The second translation-example information 300 may include information related to a number of variable portions expressed as variables 303 in the original and translated text information.

In translation-example input step 101, the first translation-example information 200 including the original text data 201 and the translated text data 202 shown in Fig. 4 is read out from the first translation-example dictionary 110.

In translation-example comparing step 102, a comparison is made to detect differing portions between the original text data 201 of the first translation-example information 200 read out in the translation-example input step 101 and an original text data of other translation-example in the first translation-example dictionary 110.

In the word class-specifying step 103, a word class of the differing portion detected in the translation-example comparing step 102 is specified.

In the variable generating step 104, the differing portion detected in the translation-example comparing step 102 and the type of the word class specified in the word class-specifying step 103 are linked so as to create the second translation-example information 300 having the original text data 301 and the translation information 302 having variable portions. Note that the variable generating step 104 may create data of the number of variable portions expressed as

variables 303 as a part of the second translation-example information as well as the original text data 301 and the translated text data 302.

In the translation-example registering step 105, the second translation-example information 300 shown in Fig. 5 which have been created in the variable generating step 104 is registered in the second translation dictionary 120.

An embodiment of the present embodiment will be described with reference to Figs. 6 to 12.

The present embodiment will be described based on an example where reference translation-example information 610 and comparison translation-example information 620 are read out from the old translation-example dictionary 401. If the reference translation-example information 610 and comparison translation-example information 620 are similar, the second translation-example information 630 including the differing portions expressed as variables are registered into a new translation-example dictionary 403.

The reference translation-example information and the comparison translation-example information are read out from the first translation-example dictionary according to steps S0 to S4 below.

A value received from the user is set as an initial value of value x to be used as a threshold value for determining whether the translation-example information are similar to each other. Variables m and n for indicating reading position of the first translation-example dictionary, y for determining the number of differing portions, and variable L used for determining the differing portion are initialized to a value 0. (Step S0).

Position m of the reference translation-example information to be read from the old translation-example dictionary 401 and position n to

be compared are determined (Step S1).

Reference translation-example information 610 is read out at an mth position from the old translation-example dictionary 401. (Step S2).

A read out position of the comparison translation-example information to be compared with the reference translation-example information 610 is set to an n+1th position. (Step S3).

The comparison translation-example information 620 is read out at an nth position from the old translation-example dictionary 401. (Step S4).

The original text data of both of the translation-example information read out are divided into a plurality of parts corresponding to units of word class and then compared so as to determine whether the translation-example information are similar to each other according to steps S5 to S10 below.

A morphological analysis of the original text 611 of the translation-example information 610 read out in step S2 is implemented and it is divided into a plurality of parts in accordance with word class units. (Step S5).

A morphological analysis of the original text 621 of the translation-example information 620 read out in step S4 is implemented and it is into a plurality of parts in accordance with word class units. (Step S6).

The comparison position of the part units divided in steps S5 and S6 is set to L+1th position. (Step S7).

A comparison is made to determine whether respective character strings of Lth parts divided in steps S5 and S6 match. Note that determination of a match may be performed using semantic correspondence. (Step S8).

If the comparison result of step S8 does not indicate a match, the number of differing position y is incremented by +1. (Step S9).

The number of differing position y is determined. If the differing position y is greater than or equal to x, it is determined that the comparison translation-example information 620 is not similar to the reference translation-example information 610, and after setting the number of differing portion y and the comparison position of part unit L is set to 0, proceed to the comparison translation-example reading process (S4) to read a new translation-example information 620 to be compared. (Step S10).

A process for expressing the differing portions of the original text data as variables is carried out in steps S31 and S32 shown in Fig. 8 and described below.

If the differing portion y is less than x, a word class of the differing portion is specified. (Step S31).

A differing portion between the original text data 611 of the reference translation-example information 610 and the original text 621 of the comparison translation-example information 620 is linked to the word class specified in step S31 and expressed as a variable. (Step S32).

A process for expressing the differing portions of the translated text data as variables is carried out in steps S33 to S38 shown in Figs. 8 and 9 and described below.

A morphological analysis of the translated text data 622 of the comparison translation-example information 620 is implemented and it is divided into a plurality of parts corresponding to word class units. (Step S33).

A comparison is made between types of the

word class specified in step S31 and the word class of the Jth part divided in step S33. (Step S34).

If there is no match between the types of word class, increment J such that J=J+1, and
5 implement comparison for the next portion. (Step S35).

If the type of word class matches, translate the Jth part of the parts divided in step S33 to the language of the original text data using
10 commonly used dictionaries 402 such as the basic dictionary or the technical term dictionary. (Step S36).

If there is no matching translation for the differing portion of the original text data and
15 the translation obtained in step S36, increment J such that J=J+1, and implement comparison for the next portion. (Step S37).

If there is a matching translation in step S36, link the Jth part of the translated text data
20 621 of the comparison translation-example information 620 and the word class of the Jth portion, and express it in as a variable. (Step S38).

A process for registering the second
25 translation-example information 630 in which a part of the translation-example information is expressed as variables into the second translation-example dictionary is carried out in step S50 shown in Fig. 10 and described below.
30 The translation-example information 630 created in steps S32 and S38 are registered in the new translation-example dictionary 403 together with the number of portions expressed as variables 631. Note that when the number of differing portion is
35 zero, it is not necessary to implement registration into the new translation-example dictionary 403. The variable L used for a process of determining the

differing portion is initialized to a value 0 (Step S50).

It is to be noted that the above embodiment has been described for a case provided

5   with a single old translation-example dictionary 401 and a single new translation-example dictionary 402, but a plurality of new translation-example dictionaries 402 may be created from a plurality of old translation-example dictionaries 401, or, a

10   single new translation-example dictionary 402 may be created from a plurality of old translation-example dictionaries 401, or a plurality of new translation-example dictionaries 402 may be created from a single old translation-example dictionary 401.

15   Also, the new translation-example dictionary 402 may not necessarily be completely new, but may be added to an existing translation-example dictionary.

An embodiment of using the number of

20   portions expressed as variables 631 is described with reference to Fig. 13. A group of original text data 720 are translated using a translation-example dictionary 710 having the number of parts expressed as variables, so as to create a group of translated

25   text data 730.

A first original text data 721 is translated into a first translated text data 731 using a first translation-example information 711 having only one portion that has been expressed as a

30   variable by a translation-example dictionary, and the thus-created first translated text data 731 is displayed in green.

A second original text data 722 is translated into a second translated text data 732

35   using a second translation-example information 712 having two parts that have been expressed as variables, and the thus-created second translated

text data 732 is displayed in yellow.

A third original text data 723 is translated into a third translated text data 733 using a third translation-example information 713

5   having three parts that have been expressed as variables, and the thus-created third translated text data 733 is displayed in red.

Further, the present invention is not limited to these embodiments, and variations and

10   modifications may be made without departing from the scope of the present invention.

The present application is based on Japanese priority application No. 2001-102266 filed on March 30, 2001, the entire contents of which are

15   hereby incorporated by reference.